International Journal of Electronics,
Communication & Instrumentation Engineering
Research and Development (IJECIERD)
ISSN(P): 2249-684X; ISSN(E): 2249-7951
Vol. 4, Issue 2, Apr 2014, 13-18

TRANS STELLAR
• Journal Publications • Research Consultancy

MULTI KEYFRAME ABSTRACTION FROM VIDEOS USING FUZZY CLASSIFIERS

PRACHI C. KALPANDE & A. S. BHIDE

Department of Electronics and Communication Engineering,

Shri Sant Gadge Baba College of Engineering and Technology, Shegaon, Maharashtra, India

ABSTRACT

With a vast number of stored videos in a video archive, it is not possible to produce by hand a trailer for every video stored. To organize and retrieve video information effectively has been a problem to be solved in the fields of database and information retrieval. The key frame extraction is a process which extracts the most representative image collections from the original video and is the basis of video analysis and retrieval. The objective of this paper is to generate multi-key frames by adding efficient Fuzzy C-Classifiers. A better and faster correlation maps can be generated to extract semantically meaningful information from the videos with overlapping views. Thus the goal is to retrieve the images meeting with specific visual feature descriptions from extensive video database, according to the features such as scenes, moving object in the video data, color, textures and shapes in the image data automatically without human involvement which facilitates quick browsing.

KEYWORDS*:* Euclidian Distance, Fuzzy Clustering, Multi Key Frame, Overlapping Views, Video Summarization

I. INTRODUCTION

A significant increase in the various multimedia applications due to the advance development in the computing and network infrastructure has lead to the wide use of digital videos. There is an amazing growth in the amount of digital video data in recent years by the use of multimedia applications in the areas of education, entertainment, business, and medicine. People are more interested in finding the information they need from the large number of video database. And therefore crucial issue is how to extract useful video content and retrieve video from huge data base.

Key frame extraction is one form of video abstraction which gives one or more salient frames summarizing the overall information of the video. Many related work has been carried out in key frame extraction, Li et al. [1] proposed a motion-focusing method to extract key frames and generate summarization for surveillance videos. Jiang and Qin [2] introduced a key frame summary using visual attention index descriptor based on visual clues model, a comprehensive summary of the state-of-art video abstractions are reviewed in [3]. In general, two basic forms of video abstraction exist, the static key frames and dynamic video skims. Ping Li, Yan wen Guo [10] propose a correlation map to naturally model the correlations with various attributes among multi-key frame, key frame importance and weighted correlations are then computed to construct the map to generate multi key frames.

This paper proposes an algorithm that utilizes Fuzzy C classifier for unique multi-key frame generation mechanism to improve quick browsing for videos with overlapping views. User can fetch the important events and receive single or multiple key frames of the events from the original Videos.

The paper is divided into sections as follows. Section 2 gives a brief explanation of Frame Extraction. A general

overview of a Frame pre-processing is given section 3. In section 4 explains some basic concepts and the general idea behind Fuzzy C-means Clustering based upon features like color, texture, content, and pattern of repetition etc. In section 5 k-Neighborhood algorithm is explained. Fuzzy c-means clustering (FCM) algorithm is presented in section 6 and Correlations of Key frames are given in section 7. Conclusion is given in section 8, references in section 9.

II. FRAME EXTRACTION

The video key frame extraction technology is one of the important part of content-based video retrieval and is the basis of video analysis and retrieval. The content based video retrieval is a process of selecting images which relates with specific visual feature descriptions from large video database, according to the features such as frames, scenes, lens, and moving object in the video data, color, textures and shapes in the image data. The key frame extraction is a process which extracts the most representative image collections from the original video.

Key frame refers to the image frame in the video sequence which is representative and able to reflect the summary of a shot content. By using the key frame we can express the main content of lens clearly, and reduce the amount of video processing data and complexity greatly. So we could make the storage, organization and retrieval of video information more convenient and efficient. That means instead of entire video information only key frames can be used to represent important information of video frames. There are many key frame extraction methods. A brief description of few of them is as follows:

2.1 Shot Boundry Method

The main concern of this approach is to detect shot boundaries of the video. The principle methodology of shot-boundary detection is to extract one or more features from the frames in a video sequence [1]. The difference between two consecutive frames is computed using the features. Colour or greyscale histograms can be also used. If difference is more than a certain threshold value, a shot boundary is declared. Once shot detection is completed, key frames are selected from each shot. The method is easy to realize and requires less computational efforts.

2.2 Image Frame Information Method

This technique uses frame averaging and histogram method. Frame averaging method calculates the average of pixel value of all t frames in a position, and then chooses the frame whose pixel value is closest to average as a key frame. Histogram method averages the statistics histograms of all the image frames and then selects the frame whose statistics histogram is close to the average statistics histogram as a key frame.

2.3 Clustering Method

It is most effective algorithm to extract key frames. The basic idea is the video frames are grouped in accordance with the correlation of the visual content by clustering, and then extract the most representative frame from each group as a key frame. The method mainly includes four functional modules; they are shot segmentation, key frame extraction, feature extraction and similarity matching. Thus clustering method groups the video frames based on the relevance of video frames and extracts most effective frames as a key frames. It reduces amount of key frames as much as possible. It is most popular method of key frame extraction now a days.

III. FRAME PREPROCESSING

Pre processing the input data is aimed at improving the effectiveness and efficiency of the clustering process. Frame Pre processing involves following steps:

- Image abstraction from videos
- Dynamic resizing
- RGB to gray conversion
- Wiener filtering
- Histogram analysis

The extracted image is resized with dynamic resizing and then converted to gray scale. The wiener filtering is used to produce an estimate of a desired or target random image by linear time-invariant filtering. The Wiener filter minimizes the mean square error between the estimated random image and the desired image. The filtering bypasses the noise that corrupted the image. Finally the filtered image is histogram equalized in order to distribute intensity equally in image as well as background.

Bounded box extraction or edge to edge detection is used in order to find discontinuities in intensity values of an image which is very useful in image segmentation. The images having salt and pepper noise is filtered using median filter.

IV. FEATURE SELECTION

The video data characteristics are generally divided into the static characteristics and dynamic characteristics. Dynamic characteristics is related to the change information of the video and is normally reflects lens feature. Key frames are extracted using static features. The common image feature extraction method uses extraction of color features, texture features, shape features and edge features.

Colour feature is usually applied to differentiate the video frames. Colour histogram is a popular method to describe the colour feature in a frame due to its simplicity and accuracy. Generally the color space includes RGB, HSV, CMY, HIS and so on. The HSV color space has been widely used because it reflects the mode of human visual observation about colors [12]. When using HSV color space, we firstly quantize the space of Colour (H), Saturation(S), Brightness (V) unequally according to the eye perception to color. In general, the hue space (H) is divided into eight parts, saturation space(S) is divided into three parts, and the brightness space (V) is divided into 3 parts. So that the HSV color space is divided into 72 sub-spaces, these three color components (H, S, V) are synthesized as one-dimensional feature vector by the following formula:

$$L=9H+3S+V$$

By calculating pixels of frame i in each subspace, we can get the HSV color histogram vector, Hi, of frame i. Then the difference between frame i and frame j can be expressed by the distance, $D(H_i, H_j)$ or Euclidean distance $D(X_i, X_j)$ of the H_i and H_j

$$D(H_i, H_j) = \sum_{k=0}^{n-1} H_i(k) - H_j(k)$$

$$D\left(X_i, X_j\right) = \left(\sum_{K=0}^{n-1}(H_i(k) - H_j(k))^2\right)^{1/2}$$

Where H_i and H_j is the HSV color histogram vector of frame i and frame j, $D\left(H_i, H_j\right)$ represent the distance of H_i and H_j; $D\left(X_i, X_j\right)$ represent the Euclidean distance of H_i and H_j.

V. K-NEIGHBOURHOOD ALGORITHM

K mean cluster is the most basic cluster method. It chooses K image frames at random as the initial cluster centers. Commonly, after video segments, the first frames of each sub - shots are the initial cluster centers. And then it calculates the distance between each image data and every cluster center, and classifies the sample into the cluster whose center is closest to it. Then it re-calculated the center of the cluster to replace the old one, then calculates classify of the next sample until all the samples are classified. The method is simple and relatively easy to achieve, but the clustering results are with a lot of randomness, and the clustering results depend on the choice of k value, which have a large limitation. When the users give a large K, it not only can determine the initial cluster centers but also can change the value of K automatically in order to choose the appropriate number of clusters. It has better adaptability.

VI. FUZZY C-MEANS CLUSTERING (FCM) ALGORITHM

Historically, the FCM clustering algorithm introduced by Bezdek is an improvement of earlier clustering methods [11]. In fuzzy clustering (also referred to as soft clustering), data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters. It is based on minimizing an objective function, with respect to fuzzy membership *U*, and set of cluster centroids *V*.

$$J_m(U,V) = \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^{m}\ d^2(X_j, V_i)$$

In the above equation, X is data matrix, *N is* the number of feature vectors, *C* is the number of clusters, U_{ij} is membership function of vector *xj* to the *ith* cluster.

$$d^2(X_j, V_i) = \left\|X_j - V_i\right\|^2$$

Is a measurement of similarity between and *xj* and *vi*

The method first considers N image frames each as a cluster, then set the distance between samples and clusters. As each image frame is considered as a cluster, the distance between clusters is the same as that of samples. Choose the two clusters which are closest and combine them as a new cluster, update the cluster center of each cluster. Then calculate the distance between the new cluster and other cluster, choose the two clusters which are closest and combine them as a new cluster. Repeat the above steps until all samples are classified into clusters or the distance is larger than the clustering termination threshold.

VII. CORRELATIONS OF KEY FRAMES

We systematically represent the correlations among key frames, and use the correlation map to naturally organize the correlation among all the key frames.

There are two kinds of correlations among multi-key frame [10]

- **Temporal Adjacency:** Two key frames have high probability of describing the same event, when one key frame is temporally contiguous to the other.
- **Semantic Similarity:** Two key frames are related to each other, when they are visually similar.

First compute the weighting coefficient of the correlation among key frames using temporal adjacency and visual similarity as:

$$w(K_i, K_j) = G(\| d_{tempo}(K_i, K_j) \|)G(\| d_{visual}(K_i, K_j) \|)$$

Where, *w* (*Ki*, *Kj*) is the weighting coefficient of the correlation among key frames; *d tempo* and *d visual* are temporal distance and visual distance between key frame *Ki* and *Kj* respectively. *G* (.) is the Gaussian weighting functions to estimate the probability of *d tempo* and *d visual* . The values of *w* (*Ki*, *Kj*) are between 0 and 1, which is property for correlation weighting functions. The covariance between *Ki* and *Kj* is computed as:

$$C_{ov}(K_i, K_j) = E\{[K_i - E(K_i)][K_j, E(K_j)]\}$$

The un weighted correlation values between two key frames are from 0 to 1, with 1 denotes two key frames are perfectly correlated, and 0 denotes no linear correlation between two key frames.

Finally, we obtain the weighted correlation among multi-key frame using the weighting coefficient and probabilistic correlation as shown:

$$Corr_w(K_i, K_j) = w(K_i, K_j)Corr(K_i, K_j)$$

The values of the weighted correlation $Corr_w(K_i, K_j)$ among frames are also between 0 and 1. Each edge of the correlation map connects a pair of nodes (key frames) with correlation evaluated using our methods based on the temporal and semantic feature similarity and probabilistic correlations among frames. Therefore, combining the correlation map structure with the key frame importance and weighted key frame correlation, we can naturally model the correlation among videos with overlapping views using our key frame correlation map.

After the construction of the correlation map of key frames, we employ the support vector machine to classify the event-centered multi-key frame on the map, and apply rough set theory (RST, one of the best attribute reduction rules) to select the most important and viable key frames in each class. A detailed description of these processes is beyond the focus of the paper. The key frames remained in all classes after RST are then assembled in temporal order for generating the final abstraction.

VIII. CONCLUSIONS

In this paper, multi-key frame abstraction is developed to efficiently browse video datasets. We propose a key frame correlation map to naturally represent the correlations among multi-key frame. The weighted correlations are computed based on probabilistic theory and the temporal and visual semantic similarity among key frames. Essential key frames are further generated for abstraction using rough set, and the event- centered correlation maps are presented to serially assemble multi-key frame along time line, to facilitate easy browsing of video datasets.

IX. REFERENCES

1. C. Li, Y. Wu, S. Yu, and T. Chen, "Motion-focusing key frame extraction and video summarization for lane surveillance system," in Proc. of the 16th IEEE ICIP, pp. 4273-4276, 2009.
2. P. Jiang and X. Qin, "Keyframe-based video summary using visual attention clues," IEEE Multimedia, vol. 17, no. 2, pp. 64-73, 2010.
3. B. Truong and S. Venkatesh, "Video abstraction: a systematic review and classification," ACM Trans. on Multimedia Computing, Communications, and Applications, 3(1), pp. 1-37, 2007.
4. C. Cotsaces, N. Nikolaidis, and I. Pitas, "Video shot detection and condensed representation: a review," IEEE Signal Processing Magazine, 23(2), pp. 28-37, 2010.
5. C. Ngo, T. Pong, and R. Chin, "Video partitioning by temporal slice coherency," IEEE Trans. on Circuits and Systems for Video Technology, 11(8), pp. 941–953, 2001.
6. A. Ferman and A. Tekalp, "Two-stage hierarchical video summary extraction to match low-level user browsing preferences," IEEE Trans. On Multimedia, 5(2), pp. 244-256, 2003.
7. L. Wu, X. Hua, N. Yu, W. Ma, and S. Li, "Flickr distance," in Proc. Of ACM Multimedia, pp. 41-40, 2008.
8. J. Deng, W. Dong, R. Socher, L. Li, K. Li, F. Li, "Imagenet: a largescale hierarchical image database," CVPR, pp. 248-255, 2009.
9. Y. Fu, Y. Guo, Y. Zhu, F. Liu, C. Song, and Z. Zhou, "Multi-view video summarization," IEEE Trans. on Multimedia, 12(7), pp. 717-729, 2010.
10. Ping Li1, Yanwen Guo2, Hanqiu Sun1," Multi-keyframe abstraction from videos", IEEE International Conference on Image Processing, pp.2473-2476, 2011.
11. Soumi Ghosh, Sanjay Kumar Dubey, "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms", International Journal of Advanced Computer Science and Applications, Vol. 4, No.4, 2013.
12. Zhonglan Wu, Pin Xu , "Research on the technology of video key frame extraction based on clustering", IEEE Fourth international conference on multimedia information networking and security,pp. 290-293, 2012.